DOCUMENT RESUME

ED 396 591 FL 023 976

AUTHOR	August, Diane, Ed.; McArthur, Edith, Ed.
TITLE	Proceedings of the Conference on Inclusion Guidelines and Accommodations for Limited English Proficient Students in the National Assessment of Educational Progress (Washington, D.C., December 5-6, 1994).
INSTITUTION	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	NCES-96-861
PUB DATE	May 96
NOTE	50p.
PUB TYPE	Reports - Descriptive (141) -- Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; Federal Legislation; Federal Regulation; Identification; *Information Needs; *Limited English Speaking; *National Surveys; Reliability; *Research Methodology; *Sampling; Testing; *Testing Programs
IDENTIFIERS	*National Assessment of Educational Progress

ABSTRACT

This report presents the results of a working meeting to provide guidance to staff at the National Center for Education Statistics on: (1) establishing guidelines for inclusion of limited-English-proficient (LEP) students in the National Assessment of Educational Progress (NAEP), field tests, research, and development; (2) modifications in the NAEP and administration procedures to make it more LEP-inclusive; (3) how to report data on LEP students; (4) major technical and implementation issues that might be part of a federal research agenda on inclusion and accommodations; and (5) monitoring and follow-up research to ensure appropriate and consistent inclusion and modification strategies. The NAEP is a congressionally-mandated assessment of what American students know and can do. An introductory section provides background on the NAEP, its purpose and legislative requirements for the data, and special considerations when including LEP children in assessment. Subsequent sections summarizes participants' discussions in each of the five areas outlined above. A conference agenda and participant list are appended. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

FL

ED 396 591

Proceedings of the Conference on Inclusion Guidelines and Accommodations for Limited English Proficient Students in the National Assessment of Educational Progress

December 5-6, 1994

Sponsored by the
National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL023976

BEST COPY AVAILABLE

NATIONAL CENTER FOR EDUCATION STATISTICS

Proceedings **May 1996**

Proceedings of the Conference on Inclusion Guidelines and Accommodations for Limited English Proficient Students in the National Assessment of Educational Progress

December 5-6, 1994

Sponsored by the
National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education

Diane August
Independent Consultant

Edith McArthur
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement **NCES 96-861**

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Jeanne E. Griffith
Acting Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue NW
Washington, DC 20208–5574

May 1996

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *Proceedings of the Conference on Inclusion Guidelines and Accommodations for Limited English Proficient Students in the National Assessment of Educational Progress*, NCES 96-861, by Diane August, independent consultant, and Edith McArthur, National Center for Education Statistics. Washington, D.C.: 1996.

Contact:
Edith McArthur
(202) 219–1442

INCLUSION GUIDELINES AND ACCOMMODATIONS FOR LIMITED ENGLISH PROFICIENT STUDENTS IN THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

EXECUTIVE SUMMARY

This document summarizes the results of a working meeting held in Washington, DC on December 5-6, 1994 to provide guidance to staff at the National Center for Education Statistics (NCES) on inclusion guidelines and accommodations for limited English proficient (LEP) students in the National Assessment of Educational Progress (NAEP).

Guidelines for the inclusion of LEP students in NAEP, fieldtests, research, and development work

Conference participants emphasized the importance of developing a set of guidelines for determining how to include students in NAEP. These would be responsive to several pressing concerns: maximizing the number of LEP students who can be validly assessed, minimizing the number of alternative testing procedures, and keeping the decision flow simple, consistent, and realistic within the NAEP context. Criteria must be developed to determine the best match between the particular characteristics of LEP students and the particular form of assessment -- an unmodified English version, a native language version, an unmodified English version with some support, a modified English version, or one of a number of alternative assessment modes.

Participants believed that only those LEP students proficient enough in English to meaningfully participate in NAEP should be given the unmodified version of NAEP. Criteria to determine ability to participate in this version of NAEP should be based on English literacy levels, rather than years in English-only instruction or other background characteristics. This is because years in English-only instruction may not accurately predict English proficiency given the tremendous variations in the home and school backgrounds among language minority students. Those LEP students who are unable to take the standard English assessment should take native language assessments if they are available and if they command the requisite levels of native language literacy. The remainder of students would be assessed using less conventional means, such as adaptations of English assessments or of assessment procedures.

Modifications in NAEP to make it more inclusive of LEP students

Possible modifications in NAEP to make it more inclusive of LEP students include developing native language versions, use of the standard English version with various types of support, and modifications of the standard English version.

It is important to consider the fact that approximately 73 percent of LEP students come from Spanish language backgrounds. For students from Spanish language backgrounds, it is realistic to develop Spanish language versions of NAEP for use with the subset of this group with

literacy skills in their native language. However, even in the development of a Spanish version of NAEP, caution is suggested by the literature on translation, including the importance of selecting appropriate translators, identifying the appropriate language for the target version of the test (given differences among Spanish spoken in different countries), identifying and minimizing cultural differences, and finding words and phrases in Spanish equivalent to those in English.

Native language assessment for students whose first language is not Spanish may not be a realistic option since native language assessments for these students may not be available in the foreseeable future given the small overall percentages of students from these language groups. Furthermore, it may be difficult for NCES to obtain a sufficient sample size under current sample designs to allow reporting test scores for each of these language groups.

Assessments in English are difficult for LEP students because they test both content concepts and language ability, particularly reading comprehension and writing. Decreasing English language load may make English language assessments more appropriate for LEP students. Alternative strategies may be divided into those that involve actual modifications of the items and instructions (simplifying the language load) and those that provide support during administration of unmodified items (i.e., providing additional clarifying information either at the end of the test booklet or throughout the text, providing taped instructions and audio tapes for answers, providing more time). In all cases, it is important to consider students' academic capability when adapting assessments.

A sizable proportion of LEP students may be left out of assessment even with the availability of Spanish assessments and these modifications. Information should be collected about these excluded students even if the data may not meet validity and reliability criteria for NAEP. For example, NAEP scores might be assigned to these students based on teacher ratings or imputation based on students language and educational background information, or some combination of these. Other alternative sources of information might include the use of portfolios, extending the concepts of scaffolding and sheltered instruction to assessment, as well as using demonstrations.

Finally, participants recommended taking into consideration the needs of LEP students during test development, such as through decreasing the English language demands of both test items and instructions. These modifications would be accomplished without compromising the validity of the assessment for English-proficient students.

Scoring

Participants stressed the importance of developing scoring rubrics and procedures that are appropriate for LEP students, i.e., that consider their linguistic and cultural background. They also recommended examining whether the imputation of scores based on student background variables was a feasible way to develop test scores for LEP students.

Fieldtesting

These inclusion strategies will require research, development, and fieldtesting before they can be implemented. Conference participants recommended that criteria be established to determine which methods can be fieldtested now and which require further research and development work. Furthermore, it will be necessary to develop guidelines for LEP student inclusion for each modification within the three categories -- unmodified English version, versions ready for fieldtesting, and versions needing further research.

Reporting data on LEP students

Participants stressed the importance of a "standardized" definition of limited English proficiency. Currently there is much variation across states and school districts in how students are identified and tested so that measures characterizing the LEP population do not reflect the same population in different jurisdictions.

Most participants recommended that for those LEP students who take the standard NAEP assessment with no accommodation, NCES report data separately on LEP students' performance, and that the data also be reported out as part of the aggregate. Ideally data would be presented in three ways: for all students, including LEP students; for LEP students only; and for all students excluding LEP students. In addition, participants felt that efforts must be made to report outcomes for other LEP students by type of accommodation.

Federal research agenda on inclusion and accommodations in assessments

Participants stressed the importance of research and development. Major research questions include:

- What is the most meaningful way to conceptualize English proficiency? What are the requisite levels of proficiency in different dimensions of English for LEP students to participate in (unmodified) English-only assessments? What are the measurement issues associated with proficiency in those dimensions?

- How are subject-matter content knowledge and English language proficiency related? What are the implications for the development of better assessments of students' content knowledge?

- What modifications can be made in large-scale assessments (both in the assessments themselves and in the procedures used to administer the assessments) to incorporate more LEP students? How do these modifications affect the reliability and validity of the assessments? How do we determine which LEP students take which assessments (by student background, language proficiency, educational history)?

- Is it possible to assign (impute) scores to LEP students based on information about their background (such as language proficiency, educational history, and academic achievement)? If so, what background variables will best predict student outcomes both on NAEP and in academic settings?

- How best can data be reported for LEP students, given methodological problems discussed in this paper?

Monitoring

It is critical to monitor the exclusion of LEP students, ensuring that all LEP students who are capable of participating do so. For school personnel (who generally make determinations on whether and how to test students), clear and unambiguous decision trees on assessment guidelines and procedures might ensure a more systematic approach to LEP student inclusion. A specific person in each district might be required to sign off for each student who is excluded and to provide additional assessment information about the student. Follow-up studies on excluded students might provide additional information about assessment procedures and modifications that might be developed or improved.

Finally, participants recommended that an advisory committee be established to provide ongoing advice to NCES on LEP student assessment issues and to review ongoing research and make recommendations on research needs.

Table of Contents

Page

Executive Summary iii

Overview 1

Background 2

National Assessment of Educational Progress 2
NAEP Exclusion Criteria for LEP Students 3
Implications for the NAEP of LEP Student Exclusion 5
LEP Student Assessment Issues 6

Principles for Developing Guidelines 9

Maximum Inclusion Principle 9
Continuum of Strategies Principle 9
Reality Principle 9
"NAEP as a Standard" Principle 10

Guidelines for Inclusion of LEP Students in NAEP and Fieldtests 11

Possible Approaches to Deciding How to Assess 12

Possible Supportive and Alternative Strategies to Make NAEP More Inclusive of LEP Students 15

Native Language Assessment in Spanish and Other Languages 15
Adaptations of Assessment Conducted in English 18
Capturing the Remainder through Unconventional Alternatives 20
Modifications during Test Development 22
Scoring 23
Methods to Ensure Comparability of Alternate Versions 24
Fieldtesting versus Research 24

Reporting Data on LEP Students 25

Major Technical and Implementation Issues that Might be Part of a Federal Research Agenda on Inclusion and Accommodations in Assessments 27

Monitoring 28

Conclusion 29

Appendix A: Working Meeting Agenda 31

Appendix B: List of Workshop Participants 35

Appendix C: List of Background Materials 43

INCLUSION GUIDELINES AND ACCOMMODATIONS FOR LIMITED ENGLISH PROFICIENT STUDENTS IN THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Report of a Working Meeting

Diane August, independent consultant
and Edith McArthur, National Center for Education Statistics

Overview

This document reports on a working meeting held in Washington, DC on December 5-6, 1994.[1]

The purpose of the meeting was to provide guidance to the National Center for Education Statistics (NCES) on:

- Guidelines for the inclusion of limited English proficient (LEP) students in the National Assessment of Educational Progress (NAEP), fieldtests, research and development;
- Modifications in the National Assessment of Educational Progress (NAEP) and administration procedures to make it more inclusive of LEP students;
- How to report data on LEP students;
- Major technical and implementation issues that might be part of a federal research agenda on inclusion and accommodations in assessments; and
- Monitoring and follow-up research to ensure appropriate and consistent inclusion and mdification strategies.

---

[1] This paper draws upon work by James Houser which describes current NCES policy on assessment of students with disabilities and limited English proficiency and proposals for research and testing of revisions to that policy. It also draws substantially, with permission, on the work of Kenji Hakuta and Guadalupe Valdes, *A Study Design to Evaluate Strategies for the Inclusion of L.E.P. Students in the NAEP State Trial Assessment*, a paper prepared for the National Academy of Education. In addition, much of the discussion about modifications to assessment can be attributed to their paper. These papers are in Appendix B. Included in Appendix A are the meeting agenda, list of participants and their biographical sketches. Appendix B also contains summaries of current research and development and background articles that report on this research.

The discussion at the conference was limited to NAEP only (including the NAEP State Assessments) and explicitly did not include state assessment programs. NAEP serves as a barometer of the educational attainment of the nation's youth. It is not used to hold districts, schools, or students accountable for performance. State assessments, on the other hand, are generally used for accountability purposes.

The paper format follows the order of the issues raised above. Throughout the report recommendations endorsed generally by the conference participants are in italics. Prior to the discussion of these issues, however, we briefly provide relevant background information related to NAEP and to language minority students.

Background

In this section information is provided about the National Assessment of Educational Progress (NAEP), its purpose and legislative requirements for the data, and about special considerations when including limited English proficient (LEP) children in assessments.

National Assessment of Educational Progress

The NAEP is a congressionally mandated assessment of what American students know and can do. It is required "to provide a fair and accurate presentation of educational achievement" (Sec. 411 of Improving America's Schools Act, PL 103-382). The NAEP is the only assessment that tests a nationally and regionally representative cross section of students at the early elementary (grade 4), middle school (grade 8) and secondary school (grade 12) levels. The law also requires that the tests be conducted in a way that ensures valid and reliable trend reporting of achievement data. NAEP test items are written to measure a well-defined content framework for each subject assessed, including reading, writing, math, science, and other areas included in the third National Educational Goal. The assessment includes multiple-choice items, as well as short and extended constructed response items.

Because NAEP collects information on how populations and subpopulations of students are performing, it is essential that the overall sample selected be unbiased. In order to ensure the representativeness of the sample, the NAEP must sample for appropriate proportions of students by race, ethnicity, sex, region, state, and community type. While NAEP does provide reliable estimates for these types of characteristics, it does not do so for limited English proficient students or for students with disabilities.

NCES has an obligation to provide information that can be generalized to represent various populations. When the data are not representative, NCES has, first, to acknowledge this fact so data users will be informed, and second, to take steps to remedy the deficiency. Section 421 (c)(3) of the *1990 Perkins (Vocational Education) Act* requires the Secretary of Education to "ensure that appropriate methodologies are used in assessments of students with limited English

proficiency and students with handicaps to ensure valid and reliable comparisons with the general student population and across program areas." NCES interprets this to apply to both vocational and non-vocational students.

Furthermore, in legislation reauthorizing the Office of Bilingual Education and Minority Languages Affairs within the Department of Education there are explicit inclusion criteria. Limited English and language minority students are to be "included in ways that are valid, reliable and fair under all standards and assessment development, conducted or funded by the Department" (Improving America's Schools Act, PL.103-382, Part F, Section 216).

NAEP Exclusion Criteria for LEP Students

Prior to 1990, NAEP procedures allowed schools to exclude sampled students if they were limited English proficient and if local school personnel judged the students incapable of meaningful participation in the assessment. Beginning with the 1990 NAEP, the NCES instructed schools to exclude students with limited English proficiency from its assessments only if all the following conditions apply:

- The student is a native speaker of a language other than English;
- The student has been enrolled in an English-speaking school for less than two years (not including bilingual education programs)[2];
- School officials judge the student to be incapable of taking the assessment.

The guidelines also state that, when in doubt, the student is to be included in the NAEP assessment.

Approximately three percent of all eighth-grade students in schools in 1992 were identified as having limited English proficiency. Approximately two-thirds of these students were excluded from the 1992 NAEP assessments. As a result, two percent of all eighth-grade students were excluded because of language barriers.[3] At the fourth-grade level in math, 75% of the LEP students sampled for participation in the 1992 NAEP were not included in the assessment because

---

[2] This provision means that a student can be excluded from the assessment if he or she has taken the subject being tested in English for less than two years.

[3] National Academy of Education Trial State Assessment: Prospects and Realities. The Third Report of the National Academy of Education Panel on the Evaluation of the NAEP Trial State Assessments: 1992 Trial State Assessments, 1993, National Academy of Education.

of their lack of English proficiency. Thus, even these guidelines resulted in the exclusion of large numbers of LEP students from NAEP. Moreover, they have resulted in differential exclusion rates across states raising questions about the validity of state-by-state comparisons.[4]

Further concerns for any discussion about inclusion of LEP students in national assessments include:[5]

- The lack of comparable state definitions of limited English proficiency;

- Current NAEP guidelines based, in large part, on length of time in English-speaking schools. Determining ability to take NAEP according to years in an English-speaking school may be too arbitrary because it is not linked to the amount of language proficiency a student may actually have. For example, some students may not gain enough English proficiency to be able to be assessed in English even though they were in an English speaking school for two years or more while others may have sufficient proficiency;

- The lack of consistent guidelines that allow local decisions to be made about the participation of students who are LEP;[6]

---

[4] For example, Texas, California, and Connecticut have high numbers of students classified as L.E.P. but differ on numbers of L.E.P. students who are excluded from NAEP. See Exclusion and Accessibility of L.E.P. Students, a report prepared by AIR for NCES.

[5] There are many similarities between factors that lead to the exclusion of L.E.P. students and those that result in the exclusion of students with disabilities. In effect, large scale assessments pose many of the same issues for L.E.P. students and students with disabilities. See Making Decisions about the Inclusion of Students with Disabilities in Large-Scale Assessment: A Report on a Working Conference to Develop Guidelines on Inclusion and Accommodations. Prepared by Ysseldyke et al, National Center for Educational Outcomes, College of Education, University of Minnesota, April, 1994.

[6] According to many meeting participants, leaving inclusion decisions up to local school personnel [school administrators, classroom teacher(s), special language teachers, school aides, or counselors] results in tremendous variation across schools in L.E.P. student inclusion in NAEP. Although all L.E.P. students in English-speaking schools for more than two years are required to be included in NAEP, a recent study on inclusion [preliminary findings from an American Institues of Research (AIR) study on exclusion of L.E.P. students from NAEP] indicates this may not be the case. Using parent or student judgement to make inclusion decisions was also ruled out because parent reports may be inaccurate and biased by parents' own English proficiency levels and students have not been reliable sources of information regarding their own ethnicity or language proficiency.

- The differential implementation of guidelines. Some students are excluded by districts and schools arbitrarily even though they meet the inclusion criteria;

- The failure to monitor the extent to which the intent of the guidelines is followed;

- The lack of accommodations or adaptations in assessment materials and procedures that would enable some LEP students to participate;

- An altruistic desire not to impose stress upon LEP students by requiring them to take an assessment they cannot fully understand because of their limited English proficiency.

Implications for the NAEP of LEP Student Inclusion

The inclusion of more limited English proficient students in NCES' studies should provide a more accurate picture of how US students as a whole are performing. For example, results for minorities may be biased because students with limited English proficiency who are excluded from NCES' surveys and assessments. This bias is more likely to occur among minority students because proportionately more of them have limited English proficiency.

Increasing inclusion also raises issues of interpretation. The value of an assessment of LEP students is questionable if it is too language dependent to be able to accurately measure content knowledge. The issue is more complex than this: including LEP students without careful construction of the assessment or accommodations may disadvantage them. But if increasing inclusion requires modifications such as the use of alternative assessments or procedures (i.e., modification of test items and support during test administration), these modified versions may not be measuring the same content as the standard assessment. Some of the modifications may result in inaccurate estimates of the ability and achievement levels of students.[7]

[7] For example, a common accommodation, providing additional time, may present validity problems in certain cases. The test scores for students who received additional time on the SAT and GRE seemed to bias the data and overpredict their postsecondary grades. (The bias equaled approximately one-third of a standard deviation.) That is, students who received additional time to take the SAT and GRE did not perform as well academically as their test scores predicted they would. Although this study does not prove that providing additional time for some students to complete NAEP would undermine its validity, it does indicate that the use of extended time, specifically, and accommodations, in general, needs to be studied carefully before being applied in NAEP.

Altering the guidelines for inclusion of LEP students may also create problems for maintaining national trend data.[8] If LEP exclusion criteria were altered for students participating in an assessment used for measuring trend, NCES could no longer make valid comparisons between years for which different criteria were used. Because the population being tested would no longer be defined by the same restrictions, measured changes in data over time could be either the result of actual changes in performance of students or the result of adding more students to the sample with limited English proficiency. However, if the criteria are changed, one solution to retaining trend data would be to retain the existing exclusion criteria for a "trend" sample. If schools have difficulty administering two different criteria for an assessment, the samples might be drawn across rather than within schools.

LEP Student Assessment Issues

Defining guidelines for LEP student inclusion in assessments is complicated by the great diversity among the LEP student population. Although most LEP students have Spanish as their language background, approximately 27 percent come from a great number of other language backgrounds. In addition to great language diversity, they come from many different language, home, and educational backgrounds. Thus, decisions about which assessment mode to use should be made for the individual student based upon that student's background characteristics .[9]
A simplistic view of LEP students, unfortunately prevalent even among educational experts, maintains the following:

> Students speak their first language (L1) at home in infancy, enter 1st grade, are served by bilingual education programs and receive instruction in L1 in grades 1 to 3 and have access to l curriculum as mainstream children. If they are exited from bilingual program and placed in English medium instruction in grade 4, they can be assessed in English at grade 4. If they are not exited and are still classified as LEP, the best language for assessment would be L1.

---

[8] NCES conducts assessments which can be used to form trend data as part of several programs. A portion of students participating in NAEP take an assessment that is designed to provide national trend data. In order to measure the trend the NAEP contains a number of test items that have not been changed over the time series. The other major national assessment trend data stem from the longitudinal studies conducted by NCES, for example the High School and Beyond Study (HS&B) of 1980 and the National Educational Longitudinal Study of 1988 (NELS:88) and the planned Early Childhood Longitudinal Study (to begin in 1997). In addition to studying assessments over time within a longitudinal study, the longitudinal data sets are sometimes compared to one another, for example tenth graders in the 1980 from HS&B and tenth graders in 1990 from the NELS. Thus, the longitudinal data sets can provide trend data as well as longitudinal data.

[9] Note, background variables are also important for imputing scores.

The reality, however, is much more complex. Even restricting discussion to grades 1 to 4, students enter all-English instructional programs or bilingual programs at different times and shift between programs. Moreover, the use of English and the native language vary tremendously from one program to another. Many LEP students in "bilingual programs" have received very little subject area instruction in their native language.[10] Thus it cannot be assumed that non-English-background children remain in the same kind of program during their entire early schooling experience (grades 1-4) and that children in bilingual programs receive most of their subject area instruction in their native language.

The situation is even more complex in grades 6-8 and 9-12, since there is generally little or no L1 instruction available. Compounding the problem is that immigrant students enter the US at all different ages so their exposure to English varies by age, length of time in the US, type of program they are enrolled in currently, and previous educational experience. Thus, LEP children in elementary, junior high or high school may include:

- Newly arrived immigrants with high literacy skills and good L1 school experiences;
- Newly arrived immigrants with low literacy skills and limited L1 school experiences;
- Students schooled exclusively in the United States and instructed in both L1 and L2 or only in L2.

Additionally, different schools offer different types of access to English. An 8th grade student schooled exclusively in English since grade 2 in a predominantly Latino urban school may, in spite of such instruction, still be very limited in his English language abilities. However, neither will he have developed his ability to use Spanish for academic purposes. For such a student, neither testing in Spanish nor in the "standard" English-version NAEP would be appropriate.

---

[10] According to a recent study by Development Associates, only 34 percent of L.E.P. students nationwide were estimated to receive intensive special services with significant use of the native language (defined as more than 50 percent of the time the native language was used in one academic subject, or more than 25 percent of the time it was used in math, science, and social studies combined). Note that "significant use of the native language" as defined for this study is still quite limited in terms of total use of native language for subject area instruction. For further details see Fleischman, H. L. and Hopstock, P. J., *Descriptive Study of Services to Limited English Proficient Students*, Arlington, Virginia: Development Associates, 1993.

Content area and domain of assessment complicate the situation even further. Some content areas being assessed are more dependent on language than are others (for example, reading versus math). Moreover, the current trend in assessment is increasingly language-based (for example requiring an explanation for a solution to a mathematics problem). While already difficult to disentangle for LEP students, increasing use of language-based assessment makes the separation between language proficiency and demonstration of content knowledge even more complex.

Principles for Developing Guidelines

Following are a series of principles to guide research and analysis which were supported by meeting participants. Consideration should be given to developing a coherent framework for inclusion based on elements of these principles.

Maximum Inclusion Principle

Ideally, every student in each state, regardless of language characteristics, should have an equal probability of being included in the assessment sample.

Continuum of Strategies Principle

Looking for a single strategy to enable LEP students to participate in NAEP is unrealistic since "one size fits all" will not work. Rather, the appropriate view is that there is a continuum of options available to support assessment, ranging from tested and proven to untested and unproven. These options should be treated as a working set, with ongoing attempts to (1) maximize the number of students who are offered options on the tested/proven end of the continuum, and (2) test and research the feasibility, operational impact, and reporting impact of options on the untested/unproven end of the continuum. Using the entire range of the continuum would enable inclusion of all students, even though some of the students would only be included through the use of non-comparable assessment strategies.

Use of supportive and alternative assessment strategies requires research, analysis, and evaluation to determine their comparability to those strategies used to measure the progress of fluent English speakers. Supportive and alternative assessment strategies include assessment in the native language for students who are more competent in that language, bilingual assessment, assessment in English using special administrations such as presence of translators to read instructions, extra time, scaffolding (e.g., providing contextual materials) and alternative assessments that might include portfolios and teacher assessments.

Reality Principle

Only options that are realistic in the context of NAEP (policy, reporting requirements and budget) should be considered. This principle would lead to the choice of group-administered over individually-administered assessments whenever possible. Because of cost (the Spanish version of the 1995 field-test is $1 million), developing native language assessments in less common languages may be infeasible. The principle further requires clear groundrules and criteria that trigger the different assessment support strategies. In addition, assessment supports and alternative assessments must take into account the fact that teachers already faced with large and demanding workloads should not be unduly further burdened. Thus, in cases of special administrations, the additional burden should be on the NAEP assessor, rather than on the teacher. Or possibly, teachers could be treated as "data collectors" rather than as "respondents" (for activities other than their response to the teacher survey) and be remunerated for their work on the NAEP.

"NAEP as a Standard" Principle

Although NAEP is not a high-stakes assessment, many state and local assessments are. Because many states and local districts look to the NAEP as a model for testing and assessment procedures, it is very important that NAEP policies regarding LEP student inclusion be considered in this context. This consideration also holds true for NAEP's coverage of content and item format. For example, as NAEP uses more constructed response items and assesses higher order skills, it is likely that states will also.

Guidelines for Inclusion of LEP Students in NAEP and Fieldtests

This section provides some discussion of ways to think about assessing LEP students, whether in the native language or in English-only testing conditions. Based on the principles described above, the task is to identify a parsimonious set of guidelines that will optimize the number of students who can be validly assessed, minimize the number of alternative testing procedures, and keep the decision flow simple and realistic within the NAEP context.

Participants stressed the need for a "standardized" definition of limited English proficiency for use in NAEP, specifically, and by the states and school districts, generally. Then the development of a set of guidelines such as mentioned above would flow from this definition. In addition, there are no guidelines for LEP student inclusion in versions other than the "standard" English version. These guidelines would help in determining whether for students, for whom the "standard" is not appropriate, should be given a native language version, a modified English version, English assessment with support, some one of a number of alternative assessment modes, or, as a last resort, no assessment (possibly in those cases a teacher appraisal of how the student would have performed would be an approach to use in those cases). These guidelines will need fieldtesting, research, and refinement.

Underlying the conference discussion of assessment approaches was a basic debate regarding the overall purpose of NAEP:

- To assess how a nationally representative sample of students performs on NAEP or

- To assess fairly and accurately what students know and can do.

While for most students the two approaches would measure the same thing, for LEP students, the approaches would measure different things. This is because LEP students would be demonstrating both content knowledge and English language proficiency. For these students, there will be different inclusion strategies depending on which purpose one espouses. Proponents of the former would assess all students, with no modifications. Proponents of the latter would only include those students for whom the assessment is a "fair and accurate (valid and reliable)" measure of a student's performance.

Possible Approaches to Deciding How to Assess

Two options considered by participants were:

1. Testing mode determined by student's English ability

In general, the conference participants felt that only those LEP students proficient enough in English to meaningfully participate in NAEP should be given the assessment in English without assistance.[11] *Ideally, the best criterion to determine ability to "meaningfully participate" in an English language assessment is English literacy level, rather than years in English-only instruction (or native-language instruction) or other background characteristics.*[12] This is because years in English-only instruction may not accurately predict English proficiency, given variations in language, home, and school backgrounds previously described.[13] And a measure of "proficiency" should not be limited to oral language proficiency because a measure of oral language is not sufficient to determine whether an LEP student can meaningfully participate in a written language assessment such as NAEP. Hence, measures used to determine how a student should be tested should measure proficiency encompassing a measure of literacy.

Proponents of this approach would recommend that LEP students, who are unable to take the English assessment, be assessed in their native language if possible. However, this decision should be made based upon native language literacy levels. Then, students for whom an English language assessment was determined to be inappropriate and for whom a native language assessment either was not available or was not appropriate would be assessed using less conventional means. For example, students near the English literacy cut-off score might benefit from English language assessments that are linguistically simplified. Students near the cut-off scores in both languages might benefit from bilingual versions of the assessment or an English version that provides an on-line glossary. Participants raised the following issues that need resolution:

---

[11] One definition of meaningful that emerged is scoring above chance.

[12] While not appropriate as part of the set of guidelines for determining whether to exclude students from the "standard" NAEP, all participants agreed that NCES should set a time limit on how long LEP students can be waived from taking the same assessments in English as their English-speaking peers. Because many states follow the lead of NAEP in this area, it would be beneficial for NCES staff to consult with states to arrive at guidelines for such an "outside" time limit.

[13] Participants did recommend research to determine if background variables could be predictive of ability to meaningfully participate in NAEP, but thought overall that student's current English literacy level would be the best predictor.

- Why should students be screened for English literacy levels? Screening is necessary because there is tremendous variation across states and local districts in the definition of limited English proficiency and thus tremendous variation in the English literacy levels of students defined as LEP or language minority. Some participants recommended screening all language minority students for literacy levels. Others recommended bringing into the decision any existing standardized test scores of language minority students.

- What screening instrument or procedure should be used to assess literacy? Language minority students might be administered a short screening test (newly developed or adapted from an already existing instrument) to determine levels of English literacy. An alternative would be to use current scores including literacy subtests of language proficiency tests or reading/language arts scores on standardized achievement tests or on other assessments. However, a problem with using students' existing test scores is that they may not be current (and hence not reflect current language ability) also they may not be able to disentangle reading versus language problems. (Perhaps thresholds might be set for existing measures of literacy and only students scoring below these thresholds would be given an individual literacy assessment prior to NAEP.)

- What level of literacy is adequate to meaningfully participate in NAEP?

- Which LEP students should take the native language assessments rather than the English versions and which should take other forms of assessment such as bilingual versions or modified English versions?[14]

An intensive research and evaluation effort will be necessary to determine appropriate criteria for including students in the initial screening and to develop a cost and time effective, as well as reliable, approach to assessing students' "NAEP-readiness" and the selection of appropriate alternative testing approaches.

Implementation of an approach which tailors NAEP testing mode to a student's English proficiency would require the development, validation, and adoption of a standard procedure to determine 1) cutoff levels of English proficiency and 2) English literacy level in order to determine whether the student should take the standard English-language NAEP.[15] In this approach, all language minority students who had ever been (or were recently) classified as LEP would be screened. The assessment would begin by evaluating English proficiency. If a student passed a certain threshold, the assessment would become one of English literacy. Again, passing

---

[14] Some participants felt that all L.E.P. students should be given the English version first, even with accommodations, before being given NAEP in their native language, if available.

[15] One way to implement this general approach that was recommended by a conference participant would be to use computer-assisted assessment.

a pre-determined threshold would send the student into NAEP. Scores for only those students who had answered a certain number of items correctly on NAEP would be used. However, data on English proficiency and literacy would be available for all LEP (LM) students assessed. This model could also be expanded to determine and possibly administer other versions of NAEP, including native language versions and modified English versions.

The clear advantage of this approach is that it would standardize the inclusion procedures and provide accurate information regarding literacy and proficiency levels for both included and excluded LEP students. Moreover, this information could be used to correlate existing standardized language proficiency assessment scores with NAEP performance and provide useful information on LEP student reclassification criteria and levels of English proficiency needed to participate in English-only instruction. If computer assisted language assessment could be developed and implemented, it would not create undue burden at the local level.

2. Testing all students using current NAEP materials (English and Spanish)

A second approach not widely supported by the conference participants would be to include all LEP students in NAEP regardless of English literacy levels. (Possibly for those students who were literate in Spanish but not English, a Spanish version of NAEP would be administered.) A strength of this approach is that it would automatically standardize the inclusion procedures and would not cost the additional time or money to assess English literacy.

A number of participants were concerned that this approach would force many LEP students to take a test they could not comprehend. It is likely that many of these students (those not literate in English or Spanish) would complete only a few items correctly. For many of these students, scores would be based largely on imputation. For these students, background variables could be used to generate (impute) their scores. NCES with collaboration of experts in the areas of assessment and LEP student education would need to determine what background variables for LEP students best predict NAEP outcomes.

Participants felt that the most significant drawback to this approach is that imputed scores based on the standard version of NAEP may or may not provide much information on what these students actually know and can do. Because of these concerns, the first option presented was the more strongly supported by the conference participants and the following strategies reflect this preference.

Possible Supportive and Alternative Strategies to make NAEP more Inclusive of LEP students

Following is a discussion of a variety of supportive and alternative strategies discussed by the conference participants to make NAEP more inclusive of LEP students. They include testing in the native language and strategies for testing in English with various types of support strategies. These strategies will require research, fieldtesting, and evaluation before they may be implemented.

Native Language Assessment in Spanish and Other Languages

Current Projects at NCES: NCES is currently developing Spanish language assessments. In 1995, NCES funded the Educational Testing Service (ETS) to implement a field test of the mathematics assessment to determine the feasibility and validity of using Spanish and/or Spanish-English bilingual versions of the NAEP for grades 4 and 8. The results will be used to determine if it would be appropriate for NCES to use a bilingual version or a Spanish-only version of mathematics questions in the 1996 NAEP. This will be determined partly by whether it is possible to scale data from a bilingual version or a Spanish-only version of the math assessment and if those results can be put on the NAEP scale. A similar field test is planned for science at grades 4 and 8 as part of the full scale 1996 NAEP.

ETS is also conducting the Puerto Rico Special Assessment Project in which NAEP math and science assessments have been administered at grades 4 and 8 in Spanish. The Spanish version was administered to a random sample of approximately 100-105 public schools, 10-15 private schools, and 7 Department of Education experimental schools at grade four and approximately the same number of schools at grade eight in Puerto Rico. ETS is currently conducting the data analyses which include item analyses and differential item functioning (DIF) analysis. In addition, they are exploring the feasibility of equating to the national data and scaling of results, but ETS believes that it is unlikely that the results will be comparable to the main NAEP.

Conference Participant Discussion: Assessment of Spanish-speaking students

Approximately 73 percent of LEP students come from Spanish language backgrounds. For students who come from Spanish language backgrounds, it is, therefore, realistic to develop an assessment in the native language. In relation to other language groups in the US, conference participants agreed that assessments in Spanish were most likely to cover the largest proportion of LEP students. However, even for a Spanish version of NAEP, many issues arise.

First, it is important to ensure that the Spanish assessment is equivalent to the English assessment. Conference participants discussed the difficulty of adapting tests to another language. Four issues were addressed:

- The selection of appropriate translators;

- Identifying the appropriate language for the target version of the test;
- Identifying and minimizing cultural differences; and
- Finding equivalent words or phrases.[16]

In the area of identifying the appropriate language, the issues are related to word frequencies in both languages and dialect. It is important to ensure that the words used in translation to a second language are as frequently used in that language as in the original. Frequency of usage of words is highly correlated to familiarity with those words. Thus, without comparing frequency of word usage between two languages, a straight translation may result in the difficulty of an item being greatly increased or decreased. This can happen if the words used, while meaning the same thing, are not comparably familiar in the two languages. While there are tables of frequency of word usage in English, there are no such tables in Spanish or many other languages. Also, more than one Spanish version of NAEP may be necessary, given the different dialects of Spanish spoken in the United States. In this case, a sufficient sample size within a randomly-drawn national sample for each version would be necessary.

Several participants suggested that an alternative to multiple translations would be including synonyms in the text and choosing vocabulary that did not vary by country of origin.

---

[16] Ron Hambleton recommends that two groups of translators work independently, translating the assessment from one language to another. After they resolve their differences, a third group verifies the accuracy of the translation by examining how the differences were resolved. He also recommends back-translation. Finally, he recommends validating the translated version with empirical evidence. By using item response theory, students' responses on the English version are compared with students' (fully proficient in the non-English language) responses on the translated version. In a second design, children who are competent in both English and Spanish, are given both versions of the test in counterbalanced order or students from each group are randomly assigned one version or the other. In both cases, item responses are compared across versions to make sure the item characteristic curves are similar. The principal advantage of the item response model approach is that the equivalence of items in English and Spanish can be studied even if there are ability differences in the English and Spanish examinee samples. In all cases, it is important to have a sample with spread in scores. (For a full discussion of these issues, see Ronald K. Hambleton, "Translating Achievement Tests for Use in Cross-National Studies", *European Journal of Psychological Assessment*, Vol. 9, 1993, Issue 1, pp. 57-68. A second reference is Linda Cook at ETS who studied how to link Spanish and English versions of SAT, using item response theory.)

Another issue is whether students should be tested in their native language or the language in which they have received specific content instruction. Most native speakers of Spanish in the United States are instructed in English (even students instructed in bilingual education programs receive much of their content instruction in English). Thus, an assessment in Spanish may not be appropriate for these students.[17]

Conference participants felt that decisions about the language for assessment should depend on how much instruction in the native language students had received in the specific content to be assessed, i.e., science and mathematics.

Other participants raised issues of predictive validity -- if students are assessed in their native language, how well will this predict their performance in that content in an English-speaking environment? Some participants suggested native language assessments of content knowledge combined with measures of English proficiency could be predictive of LEP student achievement.

However, bilingual assessment is not universally supported among practitioners. For example, a group of experts convened by the California State Department of Education wrote,

> Bilingual structured assessments, defined here to mean a single assessment instrument or procedure administered during a single time period in two languages, are extremely difficult to design and almost impossible to evaluate in any meaningful way. In most cases, such assessments are unlikely to reveal anything more informative than would be obtained from separately administered tests in two languages. Because of the problems associated with developing, administering, scoring, and interpreting results as well as financial constraints associated with mixed language assessments, their use is not recommended as a general practice for large scale assessments of language or academic matter. [18]

Conference Participant Discussion: Assessment of speakers of languages other than Spanish
About 27 percent of LEP students are speakers of languages other than Spanish. Assessments of these students pose additional problems:

- First, conference participants agreed that it is not realistic to assume that native language assessments will be available for all students, given the large number of other languages in use.

---

[17] One participant reported that in California, when students in bilingual programs were given the state assessment (CLAS) in Spanish and told to circle what they didn't understand, they circled everything.

[18] Assessing Students in Bilingual Contexts: Provisional Guidelines (p. 9). Bilingual Education Office, California State Department of Education. July, 1994 (Prepublication Edition).

- Second, so few students with limited English proficiency speak any language other than Spanish NCES would be unlikely to obtain sufficient sample size under current sample designs to allow reporting test scores for each language.

The answer for testing LEP students whose native language is other than English or Spanish, as well as for those LEP students whose language background is Spanish but whose proficiency is not strong enough to be tested in Spanish, may lie in using adaptations of English assessments.

Adaptations of assessments conducted in English

Assessments conducted in English are difficult for most LEP students because they assess both content concepts and English language ability, particularly reading comprehension and writing. The interconnection between language and content in the assessment procedure makes it difficult to isolate one feature from the other. As a result, it is difficult to know whether a student is unable to demonstrate knowledge because of a language barrier or whether the student does not know the content material being tested.

Decreasing English language load may make assessments of content conducted in English more appropriate for LEP students. The list of alternative test strategies is large, but it may be divided into those that involve actual modification of the items and those that provide support during administration of unmodified items. In all cases, it is important to consider students' academic capability when adapting assessments. For example, choice of the reading level of dictionaries would have to be driven by the age/grade level of the student.

One of the conference participants reported on a CRESST/UCLA study of the impact on results of assessments of LEP students when the English used in items was modified while leaving content at the same level of difficulty.[19] Results of this study of linguistic modification indicate that overall there is no statistically significant improvement in the performance of these LEP students. However, when the researchers split the students into three ability groups some differences appeared:

- Students in the lowest categories of math class (ESL) showed slight improvement in their math performance on the revised (linguistically simplified) items;

- Students in the intermediate categories of math class (remedial/basic, low, and average) exhibited the largest improvement; and

- For the highest-level math classes, there was no improvement.

---

[19] Abedi, J., Lord, C., and Plummer, J. (1995). Language background report. Los Angeles: UCLA Graduate School of Education, National Center for Research on Evaluation, Standards, and Student Testing.

More research and development is necessary before this technique can be used for NAEP items.

The CRESST study simplified syntax (sentence structure). As a further degrease in language loading semantic (vocabulary) simplification might also be beneficial. There was discussion about whether to simplify vocabulary directly related to the content being assessed, vocabulary less related to the content, or both. *Conference participants agreed that semantic modification, while retaining the same level of conceptual complexity, is a promising approach to explore further.*

Participants also recommended that the language used in the general test and specific items be examined and possibly modified and that test instructions be made more explicit. For example, one participant cited research that indicated the more explicit the instructions, the better females and minorities do. An example is the "Draw a Person Test" where there are many assumptions about what the test taker is supposed to draw, yet this is not clear from the instructions. When the directions are made more explicit, females and minorities perform better.

Participants recommended that experts in the assessment of LEP students work with test developers to think about ways to maintain content difficulty of test items while making the language used more comprehensible. Several participants suggested that one typical way item difficulty was increased (thus increasing discrimination at the top end of scoring) was to increase the semantic difficulty of the items.

Participants recommended that there be various versions of "simplified English" tests, corresponding to the English proficiency levels of examinees.

Modifications that might provide support during administration of unmodified items were also recommended for further research and fieldtesting. One procedure entails providing clarifying information either at the end of the test booklet or throughout the text. One format might be an English-to-native language glossary for difficult vocabulary at the end of the test booklet. A second format might provide on-line synonyms for more difficult words.[20] A second modification would provide students with taped instructions and audio tapes for their answers, thus decreasing reading and writing English language load. A third modification would be to increase test-taking time. This would be especially useful for students who are using bilingual versions or versions with dictionaries. It would also benefit LEP students who are processing an unfamiliar language and content simultaneously.[21] Many of these modification may be beneficial in the testing of other students, not just LEP students.

---

[20] Providing glossaries and on-line synonyms is difficult because of the inextricable connection between language proficiency and content knowledge. By helping with language proficiency, one might also be aiding content knowledge, thus providing students with information that is being assessed.

[21] One participant reports that many L.E.P. students do not finish the Graduate Record Exam (GRE) and thus the time limit may be a major impediment.

Capturing the Remainder through Unconventional Alternatives

A sizable proportion of the LEP student population may still be omitted from assessment activities even with the availability of Spanish assessment and some modifications.[22] Infor-mation should be collected about those excluded even if the data may not be fully valid and reliable. Several participants suggested that this student background information may be useful to assign "NAEP scores" for these students. Several methods might be used to generate such scores. One method is to assign test scores based on teacher ratings.[23] Another method is to impute scores based on background information. To properly estimate test scores based on background information, however, there must be adequate and appropriate information about different kinds of LEP students. Some participants were uncomfortable with imputed scores, given that there is very little evidence regarding which background variables best predict performance for LEP students. In addition, participants were concerned about the difficulty of collecting valid background information given cost, time, and privacy concerns.

Participants recommended that background information include information about language background, language acquisition patterns, home environment, and school environment, including duration and extent of exposure to native and English language and exposure to the content to be assessed.[24] With the 1995 field test of the 1996 NAEP assessment, a new questionnaire is being fieldtested for all identified students with disabilities and LEP students which will elicit much useful background information for both included and excluded students.

---

[22] The Prospects study (Abt Associates), in their oversampling study of L.E.P. students, offered the possibility of administering students achievement tests in math and reading using the Spanish SABE, considered roughly equivalent to their primary outcome measure in English, the CTBS. Even when this possibility was available, approximately 25% of L.E.P. students were excluded from either assessment.

[23] One way of obtaining ratings is to ask teachers knowledgeable about the students how they think the students would have performed on this test. For example, a teacher may be asked to "imagine that the student took the test today." The teacher would then be asked to assign scores to the students as if they had taken the test. Another method is to provide teachers with examples of student work or descriptions of student performance and ask the teacher to rate the student, based on these examples. It would be preferable to obtain more than one rating per student.

[24] Such variables might include: time in an English-speaking school; percent time in English language instruction; percent time in native language instruction; percent time in content instruction in the subject to be assessed; recency of native language instruction.

Participants recommended that, in the future, background questionnaires be reviewed by experts in the education of LEP students. For example, current questions that use "special language programs" as a "catch-all" for programming for L.E.P. students are not useful given the multiplicity of settings and methods in which L.E.P. students are educated. assessed be included.

Several participants recommended collecting information on the larger group of all language minority students, many of whom were formerly classified as LEP (who would currently be classified as fully English proficient (FEP)), rather than restricting the data collection to currently identified LEP students.

Alternative methods of assessing the proficiencies of LEP students were discussed. Some of them, however, may be more appropriate for state or local level assessment use. Potential alternative assessments methodologies include:

- Using portfolios to collect the student's best work over time;[25]
- Developing computer-assisted assessments that are tailored to respond to the language needs and content knowledge of individual LEP students;
- Extending the concepts of scaffolding and sheltered instruction to assessment, as well as using dynamic assessment to ascertain what learning is accessible to students in their "zone of proximal development" both with and without help;[26]
- Giving assessments that are less language dependent, such as demonstrations.

---

[25] Portfolio assessments are considered by some to be potentially more informative about a student's achievement level than a paper and pencil test. Portfolio assessments have not been used to conduct large scale assessments for statistical purposes. In NAEP, however, it has been demonstrated that methodologies can be devised that permit uniform measurement on a wide variety of student writing. If the collected work of L.E.P. students falls outside the range of work that can be uniformly measured, it will require separate reporting.

[26]Sheltered instruction and scaffolding refers to contextualizing language for students. Examples provided by Hafner include surrounding difficult vocabulary or ideas with informal definitions, repetition, paraphrasing, visual aids and realia, vocabulary building, use of literary works with predictable story structures and patterns, examples, comparisions, contrasts and similar activities. For dynamic assessment, Hafner references Spector's research in which teachers give a child hints and prompts at different levels of complexity during the assessment. Notes are made about the student's ability before and after the test. See Making Our Assessments Comprehensible to English Language Learners, Anne L. Hafner, California State University, Los Angeles, CA, October 1994.

Because of possible biases in assessment results, some participants did not support allowing the same accommodations for assessment as are used in the classroom, unless absolutely necessary. [27]

In incorporating modifications to tests or testing procedures, difficult issues related to maintenance of trend data will have to be resolved. In order to preserve the ability to present trend data, some part of NAEP and its sampled population would have to remain the same. This would involve preserving the ability to make the determination about which LEP students, under the current guidelines, would have taken the current NAEP and which would have been excluded. (This is especially difficult given the variability in current practices.)

Modifications during Test Development

A more far-reaching way to include LEP students in NAEP would be to consider them during instrument development. For example, more items with less language load might be added to NAEP to enable more LEP students to participate. This would include adding more constructed response items with "simplified English" and ensuring instructions are "linguistically straightforward". By enabling more LEP students to meaningfully participate, the reliability of the assessment might be enhanced.[28] These modifications would have to be accomplished without making the assessment invalid for non-LEP students. LEP students would have to be considered in developing the NAEP frameworks or this strategy will not work. For example, the math frameworks specify that communicating what you know is as important as what you know. This has implications for LEP students who might know the answer, but have trouble communicating it.

Another issue that would have to be addressed is how to make items conceptually more difficult without increasing the semantic difficulty. In addition, in translating language items for NAEP, if words in the items do not translate well from English, they could be modified in the English version to accommodate the Spanish (or other) language version.

---

[27] One participant pointed out that when L.E.P. students taking the California assessments (CLAS) in Spanish were read the instructions in Spanish, they were inadvertently coached by test administrators.

[28] New Standards Project, for example, found "constructed responses" were less reliable and, thus, more items are needed to make the assessment reliable. For students who do less well, the assessment becomes less reliable because fewer items are attempted.

Scoring

If imputation is used to develop test scores for LEP students, a decision needs to be made about whether nonresponse on an item because of student's limited English proficiency is counted as incorrect or missing. Analysts working with data from the 1992 National Adult Literacy Survey (NALS) had to make similar decisions on what to do when persons completing that assessment completed fewer than five items. Using information recorded by the interviewers about why the adult stopped the assessment, the analysts determined if they stopped due to literacy-related reasons or not: if due to literacy-related reasons, then nonresponse was considered to be error; if not literacy-related, then scores were imputed based upon scores of persons with similar background characteristics.

Participants recommended that research be done on whether a scoring model similar to the NALS would be feasible and appropriate for LEP students in the NAEP. Applying a similar procedure as that of the NALS, imputation of scores in the NAEP for LEP students would require both background information about students and information about how that student performed in that assessment or would have performed. As previously mentioned, however, many participants were concerned about the validity of imputed scores for LEP students given the lack of research on background variables for this population.

Further issues remain outstanding for scoring of LEP student assessment materials:

- Scoring rubrics and procedures would have to be developed to enable constructed response items to be appropriately scored for LEP students. Participants stressed the importance of developing scoring rubrics and training procedures for constructed

  response items that are sensitive to the language characteristics (separating out language proficiency from content knowledge in areas outside of English language arts) and cultural characteristics of the language minority students.

- Participants noted the importance of accurate translations of the scoring rubrics and instructions and recommended that the same procedures used for translating tests be used for translating instructions and scoring rubrics.

- Scoring guides should contain exemplars of student work at varying levels of English proficiency, for different response preferences and modes, and for dialectical variation.

- Also decisions about scoring of tests taken by LEP students need to address how to score responses in non-English languages, including responses using code-switching.

- Decisions must be made about how to "weight" English language proficiency in scoring items in both the standard NAEP and modified versions of NAEP.
- Any new administration procedures will require special training and monitoring of the test scorers.

Methods to Ensure Comparability of Alternate Versions

Whatever adaptations may be used, it is imperative to obtain independent verification of the comparability of the content of the items. In addition, exploration of systematic differences in performance between Spanish and bilingual side-by-side versions is needed, as well as between these versions and those that are in English or in English with modifications for English language learners.

One potential method for gauging the comparability of test items is through DIF analysis, examining how items behave for different groups of students. If patterns of response differ for different groups, the items might not be comparable. In the Puerto Rican study, for example, some of the translated items had flat curves and upon inspection were found to be non-comparable to the US NAEP items. Flat response curves may also indicate that students have not been exposed to the curriculum.

Fieldtesting versus Research

Criteria must be established to determine which version and methods can be fieldtested now and which require further research and development work. Currently, some procedures are being fieldtested, including Spanish and bilingual side-by-side versions of NAEP and the use of threshold literacy levels as a prerequisite for taking the standard version of NAEP. Other accommodations such as extra time might soon be ready for fieldtesting, whereas simplifying English, use of glossaries and dictionaries, computer-assisted assessment, and other modifications previously mentioned will probably need further research and development.

It will be necessary to develop guidelines for the use of each type of NAEP. For example, assessments composed of linguistically less complex items might incorporate LEP students with "basic" English proficiency, but not beginning ESL students. Fieldtesting will be needed to determine whether the guidelines enable the particular "category" of LEP student to best demonstrate their content knowledge in a particular field.

In addition, decisions must be made about where to allocate resources for research and development. One possible approach is to start with what are considered the most valid methods and move out to less proven approaches. An alternative "sandwich" approach is to conduct research at both "ends", thus developing valid approaches as well as incorporating more LEP students into assessments (assuming more experimental methods will be more inclusive).

Reporting Data on LEP Students

Decisions about policies and resource distribution are often governed by findings from national data. If LEP students are not reported as a separate category, their special needs may be ignored in decisions about resources and policies. On the other hand, to report data by LEP status, NCES would have to significantly change its sampling frames to ensure there are sufficient numbers of LEP students to be able to report results. Also it might not be sufficient to provide total counts of LEP students. To make this determination, further research on different potential LEP subgroups (e.g. language, English proficiency) is necessary. Both the research and its implications for NAEP design are costly and time-consuming propositions.

Most participants recommended that NCES report data separately on LEP students who take the standard NAEP assessment (with no accommodation), performance, but that the data also be reported out as part of the total US aggregate. Data could be presented in three ways: for all students, including LEP students; for LEP students only; and for all students excluding LEP students.

Some participants felt that including more LEP students, without reporting these students out as a separate category, would give an inaccurate estimate of the performance of the ethnic groups to which the LEP children belonged. As such, they recommended that consideration be given to nesting LEP in language minority background for reporting purposes, if possible. However, others felt that reporting out as a total LEP group was important because that was the only way possible to provide information on the performance of a nationally representative sample of LEP students.

A further concern was that reporting out LEP students as a group, without information on opportunity to learn (access to course content, for example), would give the wrong impression about the capability of LEP students or about the system that educates them.

Participants again stressed the importance of a "standardized" definition of limited English proficiency. Reporting out by LEP status would mean very little, they maintained, without an operational definition of limited English proficiency, given the tremendous variation in which LEP students are currently included in NAEP.

Because very few LEP students will take the standard NAEP assessment, there will be a biased sample of the LEP population selected on the basis of English proficiency. *Participants felt, therefore, that efforts must be made to report outcomes for other LEP students by type of accommodation.* These scores will vary depending on student background and should be reported separately since they will not be psychometrically equivalent to one another (i.e. identical scores for students who did and did not receive accommodations would not reflect identical achievement or ability levels because of differences in the difficulty of the assessments with or without the accommodation.)

Currently, NCES staff have no plans to report LEP student data separately. Because they sample first at the school level, not at the student level, sampling frames would have to be changed to accomplish this. In addition, it is unclear whether there would be enough individuals in each standard reporting category used by NAEP (such as sex, race, region) to allow reporting data by LEP status.[29]

[29] NCES does not report data for a given population if the number of individuals in the sample is below 30 (62 for NAEP) and so cell sizes for reporting this population would have to be large enough. And, even though a cell size of 30 (62 for NAEP) is sufficient to report on a given population, it might not be large enough to make statistical comparisons among averages for different groups.

Major Technical and Implementation Issues that Might be Part of a Federal Research Agenda on Inclusion and Accommodations in Assessments.

There is considerable need for research and development if LEP students are to be equitably and fully incorporated into NAEP. The list below contains the major research issues raised by the participants at the conference. Some of the research issues apply to all students, e.g., how to ensure that assessments measure more than basic skills and knowledge, yet are sufficiently reliable and valid. There are, however, certain issues that are specific to LEP students. Many of these issues have been discussed in prior sections of this paper.

- What is the most meaningful way to conceptualize English proficiency? What are the requisite levels of proficiency in different dimensions of English for LEP students to participate in (unmodified) English-only assessments? What are the measurement issues associated with the proficiency in those dimensions?

- How are subject-matter content knowledge and English language proficiency related? What are the implications for the development of better assessments of students' content knowledge?

- What modifications can be made in large-scale assessments (both in the assessments themselves and in the procedures used to administer the assessments) to incorporate more LEP students? What do these modifications do to the reliability and validity of the assessments? How do we determine which LEP students take which assessments (by student background, language proficiency, educational history)?

- Is it possible and wise to assign (impute) scores to LEP students based on information about their background (such as language proficiency, educational history, and academic achievement)? If so, what background variables will best predict student outcomes, both on NAEP and in academic settings?

- How does one meaningfully measure opportunity to learn? For example, can background variables be used in coordination with student scores to assess opportunity to learn for LEP students?

- How best can data be reported for LEP students, given methodological problems discussed in this paper?

Participants recommended reviewing former studies to find out more about which background variables are most predictive of language proficiency. The 1982 English Language Proficiency Study funded by the Department of Education and conducted by the Bureau of the Census is an example of such a study.

Participants recommended that an advisory committee be established to provide ongoing advice to NCES on LEP student assessment issues and to review ongoing research and make recommendations on research needs.

Monitoring

It is critical to monitor the exclusion of LEP students, ensuring that all LEP students who are capable of participating do so. For district personnel (who will make determinations on whether and how to test students), clear and unambiguous decision trees on assessment guidelines and procedures might ensure a more systematic approach to LEP student inclusion. A specific person in each district might be required to sign off for each student who is excluded and to provide additional assessment information about the student. Follow-up studies on excluded students might provide additional information about assessment procedures and modifications that might be developed or improved.

Conclusion

The working meeting raised many issues about how to include LEP students in the NAEP and other national assessments but provided little resolution. Clearly the conference participants felt that the most important criteria in this work was the goal of a fair and accurate assessment of what students know and can do. This has serious implications in how assessments are developed, administered, and reported. The meeting pointed to areas which will immediately benefit from further research -- such as the development of a definition which can be applied consistently across states and schools of what constitutes limited English proficiency. Once this definition is available, its implementation requires appropriate measures to determine if an individual student is LEP and secondly, such measures could be used to determine how to assess individual students. One promising avenue for these measures would be computer adaptive testing. Other areas needing further research are development and testing of modifications and adaptations to assessments for LEP students. The conference participants felt that, even for Spanish language background students, translation of assessments into Spanish was no easy panacea for LEP students because of limitations in student content knowledge and differences in proficiency in Spanish. Finally, participants recommended that an advisory committee be established to provide ongoing advice to NCES on LEP student assessment issues, to review ongoing research, and to make recommendations on research needs.

Appendix A

Working Meeting Agenda

Agenda

Inclusion of L.E.P. Students in NAEP
N.C.E.S., Room 326
December 5–6, 1994

Monday, December 5

8:30 – 9:00	Continental Breakfast
9:00 – 9:30	Welcome (Emerson Elliott) Overview of the Meeting (Kenji Hakuta) Introductions (Delia Pompa, moderator)
9:30 – 9:45	NAEP: Current format, administration procedures, and reporting (NCES Staff)
9:45 – 11:15	Current Work on L.E.P. Inclusion in Large-Scale Assessments: Presentations and Discussion (Jim Houser)
	Exclusion and Assessibility of IEP/LEP students (George Bornstedt, Fran Stancavage)
	1995 Field Test for Including L.E.P. Students in NAEP (Larry Ogle)
	Study Design to Evaluate Inclusion Strategies (Kenji Hakuta)
	Linguistic Features of NAEP test items (Jamal Abedi)
	Puerto Rican Study (John Olson)
	Proposed Studies (NCES and OBEMLA staff)
11:15 – 12:00	Basic Research Issues (Richard Durán, Kenji Hakuta)
12:00 – 12:45	Lunch
12:45 – 2:30	Including more L.E.P. students in NAEP (Charlene Rivera—overview)
	Guiding principles; Modifications in NAEP Assessments and NAEP Administration
2:30 – 3:30	Guidelines for Inclusion of LEP students in NAEP and Fieldtests (Lorraine Valdez Pierce—overview)
3:30 – 4:30	Recommendations for Analysis and Reporting Data on L.E.P. Students (María Pennock-Román—overview)
4:30– 4:45	Review Agenda for Tuesday
4:45	Adjourn

Agenda (Continued)

Inclusion of L.E.P. Students in NAEP
N.C.E.S., Room 326
December 5–6, 1994

Tuesday, December 6

8:30 – 9:00	Continental Breakfast
9:00 – 9:15	Overview of Day's Activities and Introductions
9:15– 12:15	Recap of Recommendations Discussion and Questions*
12:15 – 12:45	Lunch
12:45 – 2:15	Research and Development Needs
2:15 – 2:45	Monitoring
2:45 – 3:00	Next steps
3:00	Adjourn

* At some point in the morning, Ron Hambleton will briefly discuss his experience with international assessments.

Appendix B

List of Workshop Participants

Conference on Inclusion Guidelines and Accommodations
for LEP Students in NAEP
December 5 and 6, 1994
National Center for Education Statistics

LIST OF PARTICIPANTS

Jamal Abedi
CSE-CRESST
Graduate School of Education, UCLA
Los Angeles CA 90024-1532

Phone Number 310-206-1532

Diane August
4500 Wetherill Road
Bethesda MD 20816

Phone Number 301-229-8270

George Bornstedt
American Institutes for Research
P.O. Box 1113
Palo Alto CA 94302-1113

Phone Number 415-493-3550

Ed De Avila
5866 Harbord Drive
Oakland CA 94611

Phone Number 510-547-8328

Richard Duran
School of Education
UC Santa Barbara
Santa Barbara CA 93106-9490

Phone Number 805-893-3555

Ray Fields
NAGBE
800 North Capitol Street NW
Suite 825
Washington DC 20002-4233

Phone Number 202-357-6938

Thomas Fisher
Program Director
Student Assessment Services Section
325 W. Gaines, Room 701
Tallahassee FL 32399-0400

Phone Number 904-488-8198

Valeria Ford
Director - Student Assessment
District of Columbia Public Schools
Rabaut Administrative Building
2nd and Peabody Streets, NW
Washington DC 20011

Phone Number 202-576-6288

Kenji Hakuta
School of Education
Stanford University
Ceras Building
Stanford CA 94305

Phone Number 415-725-7454

Ron Hambleton
University of Massachusetts
Hills South
Room 152
Amherst MA 01003

Phone Number 413-545-0262

Conference on Inclusion Guidelines and Accommodations
for LEP Students in NAEP
December 5 and 6, 1994
National Center for Education Statistics

LIST OF PARTICIPANTS

Elena Izquierdo
Director Language Minority Affairs Branch
District of Columbia Public Schools
Roosevelt Administrative Building, 3rd Floor
13th and Upshur Streets, NW
Washington DC 20011

Phone Number 202-576-8850

Eugene Johnson
Director, Research
Statistics and Psychometric Research
ETS - Center for Assessment
Mail Stop 02T, Rosedale Road
Princeton NJ 08541

Phone Number 609-734-5305

Rebecca Kopriva
Dep't of Educational Research, Administration,
and Foundations
California State University, Fresno
5005 North Maple
Fresno CA 93740-0303

Phone Number 209-278-0350

Julia Lara
Council of Chief State School Officers
One Massachusetts Avenue, NW
Suite 700
Washington DC 20001

Phone Number 202-336-7042

Reynaldo Macias
304 North Almansor Street
Alhambra CA 91801-2621

Phone Number 805-893-2250

John Olson
Educational Testing Service
Mail Stop 30-E, Rosedale Road
Princeton NJ 08541

Phone Number 609-734-1987

Maria Pennock-Roman
Educational Testing Service
Mail Stop 17-R
Rosedale Road
Princeton NJ 08541-0001

Phone Number 609-921-9000

Delia Pompa
555 Donaldson
San Antonio TX 78201

Phone Number 210-733-7343

Charlene Rivera
Evaluation Assistance Center, East
1730 North Lynn Street
Arlington VA 22209

Phone Number 703-528-3588

Ramsay Selden
Council of Chief State School Officers
One Massachusetts Avenue, NW
Suite 700
Washington DC 20001

Phone Number 202-336-7010

Conference on Inclusion Guidelines and Accommodations
for LEP Students in NAEP
December 5 and 6, 1994
National Center for Education Statistics

LIST OF PARTICIPANTS

Eugene Garcia
OBEMLA
330 C Street SW
Washington DC 20202

Phone Number 202-205-5463

P. Ron Hall
National Center for Education Statistics
555 New Jersey Avenue, NW Room 400G
Washington DC 20208

Phone Number 202-219-1828

Millie Lanuaza
OBEMLA
330 C Street SW
Washington DC 20202

Phone Number 202-205-5463

David Malouf
U.S. Department of Education, Office of Special Education and Rehabilitative Services
330 C Street SW
Washington DC 20202

Phone Number 202-205-8111

Edie McArthur
National Center for Education Statistics
555 New Jersey Avenue NW
Room 519
Washington DC 20208

Phone Number 202-219-1442

Larry Ogle
National Center for Education Statistics
555 New Jersey Avenue NW
Room 517B
Washington DC 20208

Phone Number 202-219-1683

Gary Phillips
National Center for Education Statistics
555 New Jersey Avenue NW
Room 308C
Washington DC 20208

Phone Number 202-219-1761

Paul Planchon
National Center for Education Statistics
555 New Jersey Avenue, NW
Room 413B
Washington DC 20208

Phone Number 202-219-1614

John Ralph
National Center for Education Statistics
555 New Jersey Avenue, NW Room 520
Washington DC 20208

Phone Number 202-219-2270

Jeff Rodamar
Planning and Evaluation Service
US Department of Education
400 Maryland Avenue SW
Washington DC 20202

Phone Number 202-205-5046

Conference on Inclusion Guidelines and Accommodations
for LEP Students in NAEP
December 5 and 6, 1994
National Center for Education Statistics

LIST OF PARTICIPANTS

Ken Sheppard
Project Director
Center for Applied Linguistics
1118 22nd Street NW
Washington DC 20037

Phone Number 202-429-9292

Fran Stancavage
American Institutes for Research
P.O. Box 1113
Palo Alto CA 94302-1113

Phone Number 415-493-3550

Lorraine Valdez Pierce
School of Education
George Mason University
Fairfax VA 22030-4444

Phone Number 703-993-3688

Leland Yee
Commissioner
San Francisco Board of Education
1489 Dolores Street
San Francisco CA 94110

Phone Number 415-241-6493

James Ysseldyke
National Center on Educational Outcomes
University of Minnesota
350 Elliott Hall
75 East River Road
Minneapolis MN 55455

Phone Number 612-624-4014

Susan Bowers
U.S. Department of Education
Office for Civil Rights
330 C Street SW Room 5030I
Washington DC 20202

Phone Number 202-205-8635

Peggy Carr
National Center for Education Statistics
555 New Jersey Avenue NW Room 308D
Washington DC 20208

Phone Number 202-219-1576

Emerson J. Elliott
Commissioner
National Center for Education Statistics
555 New Jersey Avenue, NW Room 400F

Phone Number 202-219-1839

Kathy Ellis
Office of the Undersecretary
US Department of Education
400 Maryland Avenue SW
Washington DC 20202

Ed Fuentes
Acting Director, National Institute on Governance,
Finance, Policymaking , and Management
Office of Educational Research and Improvement
555 New Jersey Avenue, NW Room 600
Washington DC 20208

Conference on Inclusion Guidelines and Accommodations
for LEP Students in NAEP
December 5 and 6, 1994
National Center for Education Statistics

LIST OF PARTICIPANTS

Sharif Shakrani
National Center for Education Statistics
555 New Jersey Avenue NW
Room 310B
Washington DC 20208

Phone Number 202-219-1764

Ricki Takai
Planning and Evaluation Service
U.S. Department of Education
400 Maryland Avenue SW
Washington DC 20202

Phone Number 202-401-3630

Liljana Tomic
US Department of Education, Office for Civil Rights
Elem. and Sec. Ed Policy Division
330 C Street SW
Washington DC 20202

Phone Number 202-205-8734

Appendix C

List of Background Materials

Note: Copies of background materials may be requested from Edith McArthur, NCES, 555 New Jersey Avenue, NW, Washington, DC 20208-5650 or by e-mail: Edith_McArthur@ed.gov

List of Background Materials for the Working Meeting on Inclusion Guidelines and Accommodations for Limited English Proficient Students in the National Assessment of Educational Progress

James Houser, NCES, *Assessing Students with Disabilities and Limited English Proficiency*

Kenji Hakuta and Guadalupe Valdes, Stanford University, *A Study Design to Evaluate Strategies for the Inclusion of LEP Students in the NAEP State Trial Assessment*

Diane August, Washington, DC, *Considerations for Inclusion Guidelines and Accommodations for LEP Students in NAEP*

National Center for Education Statistics, *Report in Brief, NAEP 1992 Trends in Academic Progress (pp 5-7)*

Fran Stancavage, American Institutes for Research, *Exclusion and Accessibility of LEP Students in the 1994 Trial State Assessment*

Educational Testing Service, *Plans for Including More Students with Limited English Proficiency and Individual Education Plans in NAEP*

National Center for Research on Evaluation, Standards, and Student Testing (CRESST), UCLA, *NAEP and Student Language Background: Progress Report*

Educational Testing Service, *The 1994 Puerto Rico Assessment of Educational Progress: La Prueba Especial de Progreso Educativo, NAEP en Espanol*

Kenji Hakuta, Stanford University, *Perspectives from Basic Research with Specific Reference to the Inclusion of LEP Students in NAEP*